IMPROVING RELIABILITY AND REDUCING RISK BY MINIMIZING THE RATE OF DAMAGE ACCUMULATION

M.T.Todinov
School of Engineering Computing and Mathematics
Oxford Brookes University, Oxford, OX33 1HX, UK
email: mtodinov@brookes.ac.uk

ABSTRACT

The paper introduces the principle of minimized rate of damage accumulation as a domain-independent principle of reliability improvement and risk reduction. A classification is proposed of methods for reducing the rate of damage accumulation. The paper introduces the method of substitution for reducing the rate of damage accumulation. The original assembly/system is substituted with assembly/system performing the same function and based on different physical principles. Such a substitution often eliminates failure modes characterised by intensive damage accumulation.

One of the methods discussed is an optimal replacement resulting in the smallest rate of damage accumulation and maximum system reliability. A method for achieving the smallest rate of damage accumulation for a system with components logically arranged in series has been proposed for the first time. A dynamic programming algorithm for determining the optimal variation of multiple damage-inducing factors to minimize the rate of damage accumulation, has also been proposed for the first time. The paper shows that the necessary and sufficient condition for using the additivity rule for calculating the threshold of accumulated damage precipitating failure is the factorisation of the rate of damage accumulation into a function of the amount of damage and a function of the damage-inducing factor.

Keywords: reliability improvement, risk reduction, rate of damage accumulation, method of substitution, deliberate weak links

1 INTRODUCTION

For many decades, the focus of reliability research has been primarily on methods for reliability prediction and risk assessment rather than methods for reliability improvement and risk reduction. In the cases where the focus is on risk reducing methods, a common approach in selecting risk reducing methods is the *domain-specific approach*. This approach usually consists of describing the system, identifying factors, events or circumstances with the potential of causing harm, brainstorming the different scenarios by which failure may materialise and decide on measures which prevent failure or mitigate its consequences.

Despite their popularity, the physics-of-failure models (Pecht et al, 1990; Pecht, 1990) are also domain-specific and have limited validity - only in the narrow application area where they have been developed. Physics-of-failure models and root cause analyses cannot transcend the narrow domain they serve and physics-of-failure risk reducing solutions cannot normally be used to reduce risk in other domains.

Thus, risk reducing measures aimed at eliminating *hydrogen embrittlement in welds* for example, are an important step towards improving the reliability of welds. However, these risk reducing measures are *domain-specific* and cannot transcend the narrow field of welding technology. They have no validity outside this narrow domain. Risk reduction has effectively been fragmented into risk reduction in numerous specific domains each of which employs risk reduction methods specific to the domain.

In contrast, the principle of improving reliability and reducing risk by *introducing redundancy*, *eliminating a common cause* or *condition monitoring* can be applied to improve reliability and reduce risk in many unrelated domains. They transcend the original domains they originated from and are domain-independent. The domain-independent principles and methods for reliability improvement and risk reduction do not rely on reliability data or on detailed knowledge of physical mechanisms underlying possible failure modes.

The domain-independent methods for risk reduction change drastically the existing paradigm in risk reduction, based predominantly on domain-specific methods, on a feedback from testing or customers, or on statistical, data-driven approaches.

The domain-independent methods for risk reduction provide the theoretical foundation necessary for solving risk management problems in *mechanical engineering*, *materials science*, *civil engineering* and *construction*, *electronics*, *software engineering*, *chemical engineering*, *financial control*, *management*, *environmental sciences*, *logistics supply*, *economics*, etc.

The domain-independent methods for reducing risk are so powerful, that some of them (e.g. *the method of segmentation*, *the method of separation* and *the method of inversion, Todinov (2016)*) can effectively be used as self-contained tools for solving problems unrelated to risk reduction. Consequently, the present paper contributes another domain-independent principle for improving reliability and reducing risk whose essence is minimising the rate of damage accumulation.

Failures caused by accumulation of damage with mechanisms such as fatigue, corrosion, wear, creep, have been known for a long period of time.

Dasgupta and Pecht (1991) distinguished ‘overstress failure modes’ (brittle fracture, ductile fracture, yield, buckling, etc.) and ‘wear-out failure modes’ failures (fatigue, (ii) corrosion, stress-corrosion cracking, wear, creep, etc.). Overstress failures occur when load exceeds strength. If load is smaller than strength, the load has no permanent effect on the component. Conversely, wear-out failures are characterised by damage which accumulates irreversibly and does not disappear when the load is removed. Such are for example the fatigue failures, failures caused by wear, creep, corrosion and other degradation mechanisms. A common trait of degradation failure mechanisms is that failure is precipitated as soon as the accumulated damage reaches a critical threshold level (Blischke and Murthy, 2000). Reducing the rate of damage accumulation and delaying the failure mode is therefore a powerful way of improving reliability and reducing risk.

Despite that wearout failures are very common it is surprising that minimising the rate of damage accumulation has never been formulated as a domain-independent principle for risk reduction. Furthermore, no overview of mechanisms and techniques by which the rate of damage accumulation can be minimised and reliability maximised has ever been reported.

To fill this knowledge gap, solutions related to reducing risk were collected whose underlying basis was reducing the rate of damage accumulation. A certain level of abstraction was used to strip the available solutions from their specific engineering context in order to uncover the underlying act of reducing the rate of damage accumulation. From the body of available solutions, various mechanisms emerged which were captured and distilled into distinct categories, classes and individual techniques. A classification summarising these categories, classes and techniques has been presented in Figure 1. In what follows, an extended discussion is provided on a number of key methods for reducing the rate of damage accumulation.

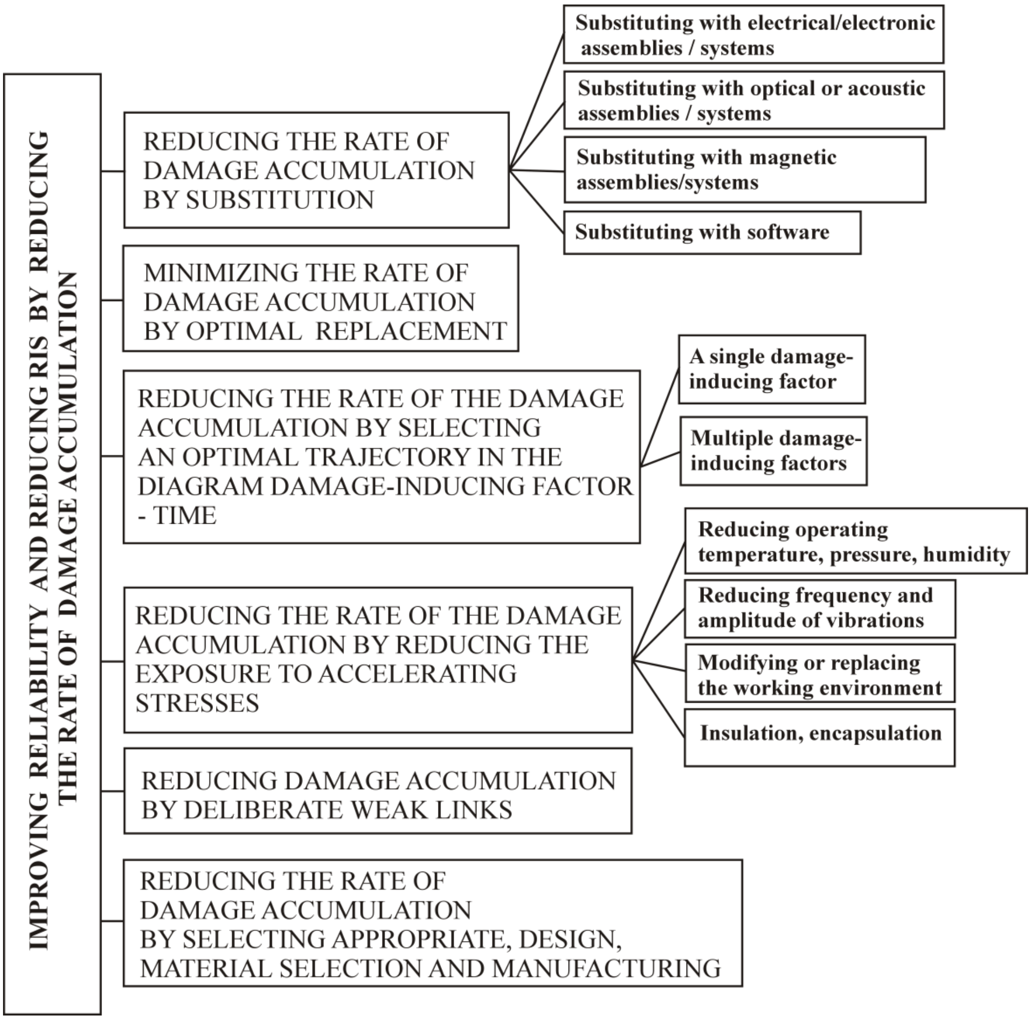

Figure 1. Classification of various domain-independent methods for reducing risk by reducing the rate of damage accumulation

1. REDUCING THE RATE OF DAMAGE ACCUMULATION BY SUBSTITUTION

A device or assembly performing a particular function can be built on different physical principles. Thus, a device measuring length could be built on mechanical, electrical, electro-mechanical, magnetic, optical or acoustic physical principles. It could also incorporate a software component. Consider a device/assembly performing a particular function in particular working conditions (environment). The device/assembly is substituted with a device/assembly built on different physical principles but performing the same function. The mechanism behind this reliability improvement method is *to eliminate or mitigate failure modes associated with intensive damage accumulation by a substitution/replacement with assembly/system working on a different physical principle*. The failure modes characterising the substituting device are associated with less intensive damage accumulation.

The argument about which of several competing assemblies performing the same function is associated with a smaller rate of damage accumulation: the mechanical, magnetic, electrical, acoustic or optical assembly *depends largely on the specific application conditions and requirements*.

In substituting mechanical assemblies with electrical, magnetic and optical assemblies and software, the failure modes of the substituting assembly need to be assessed carefully. If a substitution of a mechanical assembly introduces new, dangerous failure modes, such a substitution cannot be justified. Thus, if a mechanical system is substituted by a combination of electronics and software, all failure modes of the electronic component and the software component should be considered carefully. If the electronic component is subjected to high temperatures and vibrations or the software component has been developed quickly, by inexperienced developers using inappropriate algorithms, insufficiently tested, the substituting assembly could exhibit more dangerous failure modes compared to the mechanical system it replaces. Bugs in the software component for example, could cause failures that cannot be predicted.

Thus, solid-state relays accumulate less damage compared to mechanical relays in applications characterised by a very frequent switching, high vibration environments, dusty and humid environments. However, the mechanical relays accumulate less damage in cases of frequent current surges, voltage spikes and currents whose magnitudes vary widely from very small to very large.

Mechanical assemblies and optical assemblies have an advantage to electronic assemblies in environments characterised by a high temperature and high levels of nuclear irradiation where the reliability of the electronic circuits will be compromised.

Commonly, the substitution of mechanical assemblies with electrical, magnetic or optical assemblies and software eliminates intensive wear, fatigue, intensive corrosion and material degradation, which are major factors causing damage accumulation and wearout failures of mechanical equipment. Despite that electronic components also suffer wear, fatigue and overstress failure modes which are typical for mechanical components, the percentage of these failure modes is significantly smaller compared to the mechanical components. In addition, the quality of manufacturing of electronic components today is very high: typically of the order of 10 per million for components such as integrated circuits (O'Connor, 2002). This is confirmed by a clear evolutionary trend in many systems which started as mechanical systems and were gradually replaced by electronic, optical or magnetic systems. The evolution of the mechanical control of the air-fuel mixture and the ignition timing of car engines to *engine control unit* (ECU) is a common example. The evolution of the mechanical watches to quartz watches, the electromechanical mouse to an optical mouse, the washing machines and air conditioning units to intelligent washing machines and air conditioning units are other common examples.

Complex mechanisms can be successfully replaced by a combination of servomotors and software. In this way the complexity needed to guarantee the required kinematics is transferred from the mechanism to the software. The result is significantly simplified mechanical assemblies, with smaller inertia forces, fewer possibilities for jamming, wear and misalignment.

Software components guarantee flexibility and do not exhibit deterioration which is a major contributing factor to unreliability. Furthermore, replicating the software does not result in manufacturing variability of the software component.

The substitution by electrical and software components permits the introduction of sensing capabilities. These make the systems able to reset their goals autonomously and better adapt under changing external environment which significantly enhances their resilience. This also enhances the functionality of mechanical systems and enables them to meet a broad spectrum of user requirements. For example, some of the modern air conditioning units are capable of sensing both temperature and humidity and adapt their function to the environment through fuzzy logic reasoning.

Software units processing signals from sensors and operating actuators often substitutes control systems working on purely mechanical principle. The advantages are (i) less damage accumulation (ii) greater flexibility; (iii) better diagnostics of the problem; (iv) better adaptation to changing environmental and operating conditions; (v) more precise and adequate control. These manifest into an increase of the overall reliability. An example of such solution is the programmable engine control module (ECM) in modern cars which controls the valve timing, ignition timing, transient fuelling, air/fuel ratio, the optimal amount of fuel injected in the engine at different combinations of RPM and throttle position, water temperature correction when the engine is cold, etc. Before the engine control modules these parameters were controlled by mechanical or pneumatic units.

Here, it needs to be pointed out however, that often, software components are developed by inexperienced developers using inappropriate design, data structures and algorithms. In other cases, the software components are developed quickly without a formal verification or testing. In these cases, the substituting assembly incorporating software with bugs, could exhibit more failure modes than the mechanical or electro-mechanical system it replaces. In addition, one of the main advantages of the software components - lack of variability after copying into other systems, transforms into common cause failures due to the common software bugs.

More often, the substitution of complex mechanical assemblies with electrical, magnetic, optical, acoustic assemblies and software reduces the complexity of design, the number of moving parts, wear and fatigue and increases precision. In addition, replacing mechanical assemblies with electrical, optical or acoustic assemblies often improves maintainability which results in reduced downtime and increased availability of the device. Thus, replacing a mechanical measuring system with magnetic or optical measuring system often eliminates the need for calibration and lubrication which are necessary for conducting an accurate measurement. For example, measuring the thickness of the tube in continuous tube production by using radiography is superior to measuring the thickness by using mechanical devices. The radiographic measurement is continuous and does not require stopping the production line. In addition, it is more precise than a measurement conducted by using a mechanical device.

A good example of replacing of a mechanical assembly with magnetic assembly is provided by the magnetic stirrer. Magnetic stirrers rely on a rotating magnetic field and a stir bar to stir chemically active liquids in hermetically closed vessels. Magnetic stirrers eliminate rotating seals needed for conventional stirring and the associated with them failure modes.

Another example of replacing a mechanical assembly with magnetic assembly is provided by the magnetic worm drive (featured in the US patent US3814962, 1971) whose worm gear is made of permanent magnet material. The teeth of the worm gear and the worm wheel are also magnetised so that the like poles on the wheel and on the worm gear face one another. Magnetic repulsion transmits force from the rotating worm gear to the worm wheel which causes the rotation of the worm wheel.

The advantage of the magnetic worm drive compared to the conventional mechanical worm drive is the frictionless transfer of torque which eliminates contact stresses and wear. Wear is a major failure mode characterising the mechanical worm drive. The need for lubrication is also eliminated together with its failure modes (wrong oil used, oil degradation, oil contamination, low quantity of oil). The lack of lubrication simplifies the system and increases its reliability. Furthermore, the clearance between the teeth of the worm gear and the worm wheel eliminates failure modes caused by misalignment which enhances the life of the bearings. The clearance also eliminates the spread of vibrations through the worm wheel which reduces wear and further enhances the reliability of the assembly. Finally, the lack of lubrication simplifies maintenance which improves availability.

An example related to reducing the rate of damage accumulation and increasing reliability by replacing a mechanical component with an electrical component can be given with the mechanical push button switch. The reliability of a switch is measured by the number of actuations that can be done by the switch before an unacceptable deviation from the required performance is obtained due to damage accumulation. The damage accumulation is due to the presence of mechanical contact which promotes: (i) mechanical deterioration caused by fatigue and wear and (ii) contact erosion due to spattering and fusing of contact material caused by arcing. Consequently, eliminating the mechanical contact has the potential to eliminate the listed failure modes.

Eliminating the mechanical contact could, for example, be achieved by substituting a mechanical push button switch with a switch whose operation is based on the Hall effect, increases the durability of the switch from tens of thousands to tens of millions actuations. The Hall effect is the potential difference appearing on a conductor carrying electrical current and placed in a magnetic field. The potential difference is due to the asymmetric distribution of charge caused by the Lorenz force which is experienced by charges moving in a magnetic field.

Another example of improving reliability by eliminating mechanical contact could be given with the replacement of the contact measurement of the temperature of metal surfaces with optical (contactless) measurement by using infrared thermometers (pyrometers) (Childs, 2001). The advantages of eliminating the mechanical contact in temperature measurements are numerous:

- By using infrared technology, measurements can be made at temperatures (greater than 1300 °C). At these temperatures, even if contact thermometers were available, they would have a very limited life while the reliability of the optical (infrared) thermometers is unaffected.
- By using infrared technology, temperature measurements on hazardous surfaces can be made (for example, on high-voltage surfaces) for which contact temperature measurements are highly problematic and unreliable.
- No interference is present during optical measurements (no energy is lost from the surface during the measurement) compared to contact measurements. Because of the lack of distortion of the measured temperature, the measurement is more accurate compared to a contact measurement.
- By eliminating the mechanical contact, reliable measurements can be made at a very high speed.
- By eliminating the mechanical contact, no damage is done on the surface whose temperature is measured, which cannot be said about measurements done by welding thermocouples on the metal surface.

Fibre optic sensors technology is another area where the rate of damage accumulation is greatly reduced by replacing electrical sensors with optical sensors. Compared to electrical sensors, optical sensors are better suited for harsh environments characterised by high humidity, high pressure, high voltage and extreme temperature variations. Fibre optic sensors accumulate significantly less damage compared to electrical sensors and are better suited for a continuous condition monitoring in composite materials experiencing strain from loading or thermal expansion.

Replacing mechanical strain gauges with optical strain gauges improves the reliability of measurement because it eliminates variability due to the preparation of the surface necessary to operate the mechanical strain gauges. The variability due to the preparation of the surface leads to poor repeatability and poor reliability of the measured strain. The optical strain gauges do not require physical contact, therefore the variability associated with the different properties of the physical contact are significantly reduced.

3. MINIMISING THE RATE OF DAMAGE ACCUMULATION BY OPTIMAL REPLACEMENT

Suppose that the probability of failure in the elementary time interval $t, t+\Delta t$, given that the component has survived time *t* is given by $h(t)\Delta t$, where $h(t)$ is referred to as *hazard rate* (Fig.1a)
The reliability, associated with a time interval (0,*r*) is then given by the expression

$$R(t) = \exp\left(-\int_0^t h(\nu)d\nu\right) \tag{1}$$

where ν is a dummy integration variable. The integral $H(t) = \int_0^t h(\nu)d\nu$ in equation (1) is also known as *cumulative hazard rate*. Using the cumulative hazard rate, reliability can be presented as (Barlow & Proschan, 1975)

$$R(t) = \exp(-H(t)) \tag{2}$$

Reliability $R(t)$ can be increased by decreasing the hazard rate $h(t)$, which decreases the value of the cumulative hazard rate $H(t)$. Correspondingly, the cumulative distribution of the time to failure becomes

$$F(t) = 1 - \exp(-H(t)) \tag{3}$$

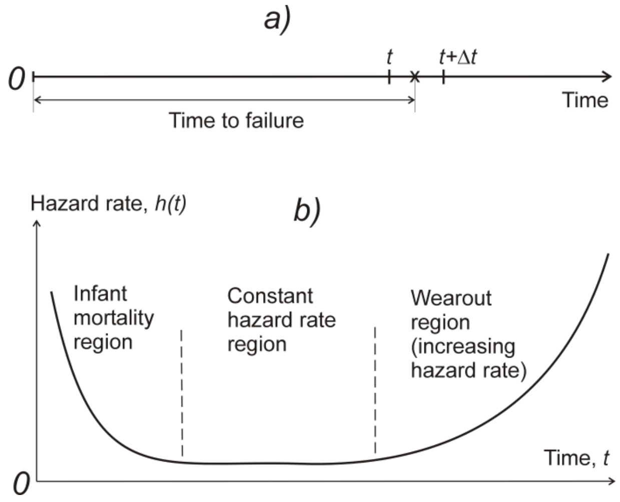

Figure 1. a) Time to failure in the small time interval $t, t+\Delta t$; b) Reliability bathtub curve

The accumulation of damage is characterised by the hazard rate $h(t)$ which normally increases with time (age). If the hazard rate practically does not depend on age, it remains constant, ($h(t) = \lambda = const$) and the time to failure distribution (3) transforms into the negative exponential distribution $F(t) = 1 - \exp(-\lambda t)$. Indeed, from $h(t) = \lambda = const.$, the cumulative hazard rate becomes $H(t) = \int_0^t \lambda d\nu = \lambda t$.

The negative exponential distribution is the model of the times to failure in the constant hazard rate region (Fig.1b). Failures in this region are not due to age, wear-out or degradation; they are due to random causes. This is why preventive maintenance in this region has no effect on the hazard rate and the time to failure distribution.

The *wearout region* of the bathtub curve in Fig.1b is characterised by an increasing with age hazard rate due to accumulated damage and degradation of properties due to wear, erosion, corrosion, fatigue, creep, etc. Unlike the preventive maintenance in the constant hazard rate region, *preventive maintenance in the wearout region significantly enhances the life of the system*.

This can be demonstrated by considering that for components experiencing a non-constant hazard rate $h(t)$, the integral $S = \int_0^a h(t)dt$ gives the expected number of failures in the finite time interval (0,a) which is numerically equal to the area S beneath the hazard rate curve (Fig.2a). Reliability in the interval (0,a) is given by $R(a) = \exp(-S)$. Consequently, replacing a component in the wearout region at time t^* (Fig.2b) reduces the expected number of failures from S to S', where S' is the hatched area beneath the combined hazard rate before and after the component replacement (Fig.2b). Consequently, a component replacement results in an increase of the probability of surviving the time interval (0,a) from $R(a) = \exp(-S)$ to $R'(a) = \exp(-S')$.

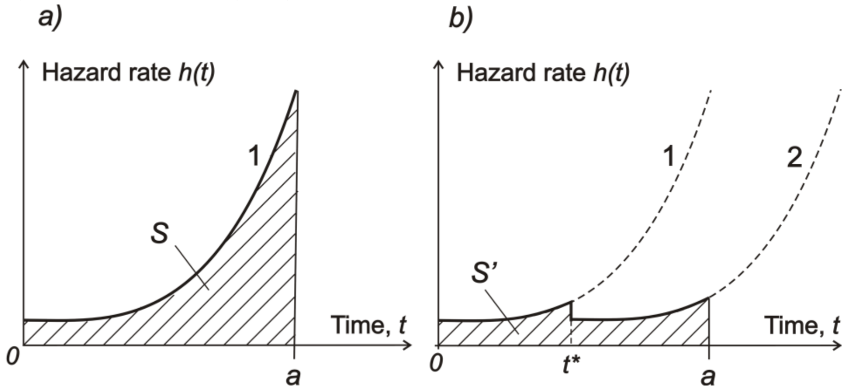

Figure 2. a) The area beneath the hazard rate curve is numerically equal to the expected number of failures in the interval (0,a). b) Decreasing the hazard rate in the wearout region by a component replacement reduces the expected number of failures in the interval (0,a).

In the wearout region, reliability is increased significantly by preventive maintenance consisting of replacing old components. This effectively delays the severe wearout phase and, as a result, reliability is increased (Fig.2b).

Consider now a time interval (0,a) and a system consisting of a number of components logically arranged in series which undergo fast wearout. A failure of any of the components constitutes a system failure. A certain number of spare components is kept for each of the working components.

Sudden system failures are highly undesirable because they cause sudden and uncontrolled shutdown which is dangerous for the system and the recovery of the system is associated with large costs. In contrast, a controlled replacement of any of the working components with a spare component can be done without disrupting the work of the system.

The question of interest is how to find the optimal replacement times which maximize the probability of surviving the operational interval (0,*a)*.

Consider a system which consists of two components logically arranged in series with a single spare component available for each of the working components. Because each component can be replaced at any time during the time interval (0,a), a combination of replacement times for the components can be represented by a point in the domain D with side a (Fig.3).

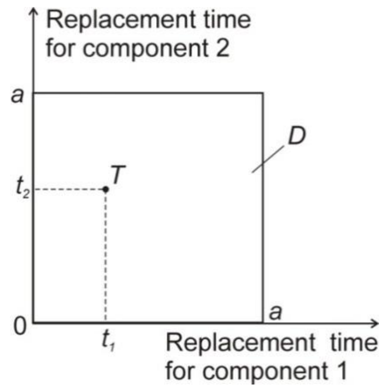

Figure 3. All possible combinations t_1, t_2 of replacement times for the two components building the system can be represented by the points from the square domain *D*.

For a specified point *T* from the domain *D*, which defines the replacement times for the components, the probability of a system failure can be determined by considering the following:

- The replacement times are always within the interval (0,*a*).
- A system failure occurs if the time to failure of any of the two components is within the interval (0,*a*)
- A system failure also occurs if after a replacement of a failed component within the time interval (0,*a*), the replacing component also fails within the interval (0,*a*).

Consider a system with components logically arranged in series, which undergo intensive wearout, with hazard rate functions strictly increasing with time. Assume also that the derivative to the hazard rate functions characterizing the separate components exists at any point in the time interval (0,*a*).

Theorem. *For an operational time interval with length 'a', if n spare parts are available for each component building the system, the optimal replacement intervals* $a/(n+1)$ *for each component minimize the rate of damage accumulation and maximize the probability that the system will survive the operational time interval.*

Proof.
The case related to *n*=1 spares for each component, will be proved first.
Indeed, because of the strictly increasing hazard rate function $h(t)$ characterizing each component, it can be shown by a differentiation that a replacement at *x* = *a*/2 at half of the operational interval (0,*a*) minimizes the expected number of failures $H(x)$ of the component during the time interval (0,*a*). The expected number of failures $H(x)$ is given by

$$H(x) = \int_0^x h(t)dt + \int_0^{a-x} h(t)dt, \quad 0 \le x \le a \tag{5}$$

Differentiating $H(x)$ with respect to the unknown replacement time *x* gives

$$\frac{d}{dx}H(x) = h(x) - h(a-x)$$

The local extrema can be obtained by equating this expression to zero:

$$h(x) - h(a-x) = 0 \qquad (6$$

Because $h(x)$ is a strictly increasing function of *x*, $h(x) - h(a-x) = 0$ only if $x = a/2$. This value is a local minimum because the second derivative, at $x = a/2$, is positive:

$$\frac{d^2}{dx^2}H(x)|_{x=a/2} = h'(a/2) + h'(a/2) > 0,$$

because $h(x)$ is a strictly increasing function and therefore $h'(a/2) > 0$. At the local minimum, $x = a/2$, the value of the function $H(x)$ is $H(a/2) = 2\int_0^{a/2} h(t)dt$.

At the ends of the interval $0 \le x \le a$, $H(0) = H(a) = \int_0^a h(t)dt$. This value is larger than the value $2\int_0^{a/2} h(t)dt$ corresponding to the local minimum because $h(x)$ is a strictly increasing function of *x*. Furthermore, there are no points in the interval $0 \le x \le a$ where the first derivative is not defined. Consequently, the local minimum at $x = a/2$ is also a global minimum

Consequently, at $x=a/2$ $H(x)$ has a global minimum which minimises the expected number of failures in the interval (0,*a*) and maximizes the probability of surviving the operational interval (0,*a*). As a result, for *n*=1, the statement of the theorem has been proved.

Now, assume that the statement is valid for any arbitrary $n-1 \ge 1$. In other words, assume that for $n-1$ spare parts available for each component, the optimal replacement times are at intervals a/n (inductive hypothesis).

Let the points 1,2,...,*n* in Fig.4a define the optimal replacement times which deliver the smallest expected number of failures over the time interval (0,*a*). Let *x* denote the length of the time interval from the start 0 of the operational interval (0,*a*) to the time of the first replacement. Let *y* denote the length of the time interval from the last replacement time to the end of the operational interval (0,*a*). If the smaller interval $(0, a-y)$ is considered (Fig.4b), the *n*-1 replacement times within this shorter interval must also provide an optimal replacement. Otherwise, the replacement times inside the shorter time interval (0, $a-y$) could be rearranged to yield a smaller expected number of failures. This smaller expected number of failures, together with the expected number of failures over the length *y* of the last time interval, will yield a smaller overall expected number of failures over the interval (0,*a*) which contradicts the assumption that the initial replacement times on the interval (0,a) are characterized by the smallest possible expected number of failures.

According to the inductive hypothesis, if the *n* replacement times in the shorter time interval $(0, a-y)$ are the optimal replacement times, they must divide the interval into equal-length time sub-intervals. In other words, for the length of the sub-interval *x*:

$$x = \frac{a-y}{n} \tag{7}$$

must be fulfilled.

Now consider the smaller interval $(1, a)$ (Fig.4c), where the *n*-1 replacement times 2,3,...,*n* within this shorter interval must also be optimal replacement times. Otherwise, the replacement times inside the shorter time interval $(1, a)$ could be rearranged to yield a smaller expected number of failures. According to the inductive hypothesis, if the replacement times in the shorter time interval $(1, a)$ are the optimal replacement times, they must divide the interval $(1, a)$ into equal-length time sub-intervals. In other words, for the length of the last sub-interval with length *y*:

$$y = \frac{a-x}{n} \tag{8}$$

must be fulfilled.
Dividing the equations (7) and (8) yields

$$\frac{x}{y} = \frac{a-y}{a-x}$$

which is equivalent to

$$(x-y)[x+y-a]=0 \qquad (9)$$

Since $x+y-a<0$, $x=y$ must be fulfilled. From $x=\frac{a-x}{n}$, $x=a/(n+1)$. Since x must be equal to all sub-intervals from 1 to n-1 and also equal to *y*, all subintervals defined by the replacement times 1,2,...,*n* must be equal to $a/(n+1)$. The theorem has been proved.

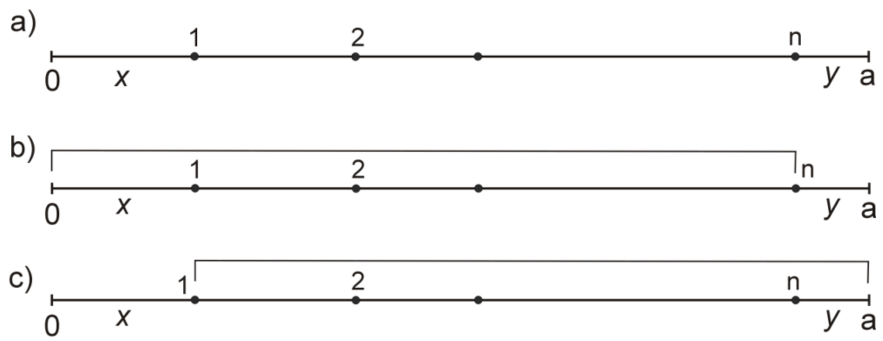

Figure 4. The optimal replacement times for a component with *n* spares should be at equal intervals $a/(n+1)$.

Example. Minimising the rate of damage accumulation by optimal replacement
Given is an operation interval with length $a=2.5$ years. The system contains two components logically arranged in series with times to failure following the Weibull distributions (with increasing hazard rate):

$$F_1(t)=1-\exp[-(t/eta_1)^{m1}];\ m_1=3.6;\ eta_1=3$$

and

$$F_2(t)=1-\exp[-(t/eta_2)^{m2}];\ m_2=1.8;\ eta_2=2.7,$$

A single spare is available for each component.
According to the proved theorem, the replacement time *t* for each of the components building the system is equal to half of the operation interval $t=a/2=2.5/2=1.25$.

This result has been verified by a simulation algorithm for determining the optimal replacement times, whose details have been omitted. The procedure for global optimization yielded an optimal replacement time of 1.25 years for each component which confirmed the theoretical result. The probability of system failure within 2.5 years corresponding to these replacement times is 0.44.

4 MINIMISING THE RATE OF DAMAGE ACCUMULATION BY SELECTING THE OPTIMAL VARIATION OF THE DAMAGE INDUCING FACTORS

4.1 A single damage-inducing factor

Suppose that the damage accumulation from a particular degradation mechanism is a function of time and a particular damage-inducing factor *p*. In the case of fatigue loading for example, the damage-inducing factor can be the stress amplitude. In the case of excessive corrosion, the damage inducing factor could be 'temperature'. In the case of wear, the damage inducing factor could be 'sliding speed', etc.

Suppose that damage is accumulated at M different intensity levels $p_1,...,p_M$ of the damage-inducing factor p (Figure 5). At each intensity level p_i, the component is exposed to damage accumulation for time Δt_i. Suppose that t_i, corresponding to constant intensity levels p_i of the damage inducing factor p, denote the times for attaining the critical level of damage D_c precipitating failure, after which the component is considered to have failed (Figure 5). It is also assumed that the sequence in which the various levels of the factor p are imposed does not affect the component's life.

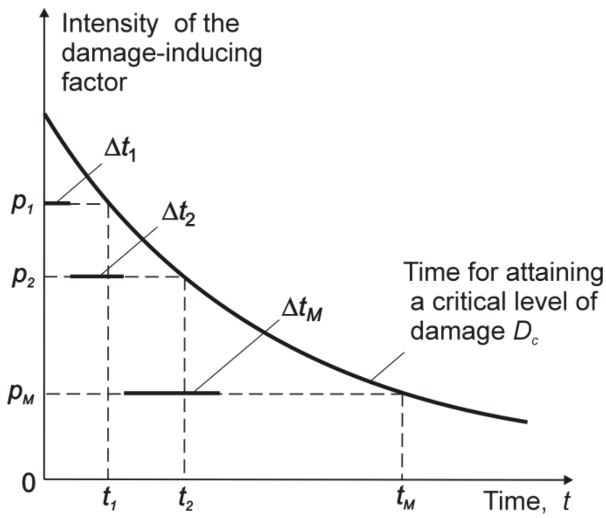

Figure 5. Exposure for times Δt_i at different intensity levels p_i, of the damage inducing factor p.

Damage factorisation is present if, for a constant level p of the controlling factor, *the rate of damage accumulation* dD/dt *can be factorised as a function* $F(D)$ *of the current accumulated damage 'D' and a function* $G(p)$ *of the damage-inducing factor* p,

$$dD/dt = F(D)\,G(p) \tag{10}$$

The critical level of damage D_c at different levels of the damage-inducing factor will be attained when the sum

$$\frac{\Delta t_1}{t_1} + \frac{\Delta t_2}{t_2} + ... \tag{11}$$

becomes unity for some k:

$$\frac{\Delta t_1}{t_1} + \frac{\Delta t_2}{t_2} + ... + \frac{\Delta t_k}{t_k} = 1 \tag{12}$$

The time t_c to attain the critical level of damage D_c is then equal to

$$t_c = \Delta t_1 + \Delta t_2 + ... + \Delta t_k \tag{13}$$

Conversely, if the time for obtaining the critical level of damage D_c can be determined using the additivity rule (12), the factorisation (10) must necessarily hold. *In other words, the damage factorisation law (10) is a necessary and sufficient condition for the additivity rule* (12). This also means that if the rate of damage accumulation cannot be factorised, the additivity rule (12) is not valid and must not be used.

Effectively, according to the additivity rule (12), the total time t_c required to attain a specified level of damage D_c is obtained by adding the absolute durations Δt_i (equation 13) spent at each intensity level *i* of the damage-inducing factor *p*, until the sum of the relative durations $\Delta t_i / t_i$ becomes unity. The fraction $\Delta t_i / t_i$ of accumulated damage at a particular intensity level p_i of the damage-inducing factor *p* is the ratio of the time Δt_i spent at level p_i and the total time t_i at level p_i needed to attain the level D_c of damage precipitating failure.

Ideally, to minimise the rate of damage accumulation and maximise reliability, the variation (trajectory) of the damage-inducing factor should be selected in such a way that *a minimum possible time is spent at the levels of the damage-inducing factor where the damage accumulating rate is high. Most of the time should be spent at the levels of the damage-inducing factor where the damage accumulating rate is low*.

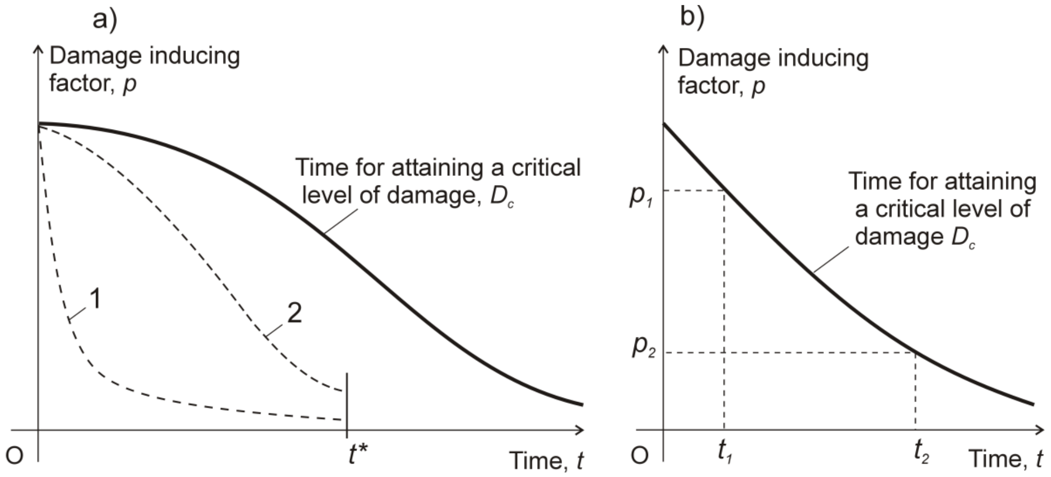

Figure 6.a) Different trajectories result in different rates of damage accumulation b) Time to failure for different intensity levels of the damage-inducing factor.

It is a well-known fact that stainless steels, with chromium added for corrosion resistance, when cooled from high temperature a precipitation of chromium carbide at the grain boundaries occurs, resulting in chromium-depleted zones along the grain boundaries. Most commonly, this occurs during welding thermal cycles and the process is known as *sensitization*. If such a weld is exposed to a corrosive environment, the chromium depleted zones along the grain boundaries are preferentially attacked and the fracture toughness of the weld is severely compromised.

Now consider the two cooling cycles of a chromium stainless steel weld in Fig.6a. The first cooling cycle ‘1’ corresponds to a fast cooling of the weld and the second cooling cycle ‘2’ corresponds to a normal cooling of the weld (2).

For the two cooling curves (trajectories) ‘1’ and ‘2’ in Fig.6a, conducted for the same length of time t^*, cooling curve ‘1’ is preferable because less time is spent at temperatures where the rate of precipitation of chromium carbides is high. Therefore curve ‘1’ is associated with a smaller amount of chromium carbides along the grain boundaries (less damage accumulated along the grain boundaries).

An important application of the additivity rule is the evaluation of the amount of damage induced by fatigue cycles. In this case, the damage-inducing factor is the loading stress range and a measure of the extent of damage is the length *a* of the fatigue crack. The additivity rule (12), also known as the ***Palmgren-Miner rule***, has been proposed as an empirical rule in case of damage due to fatigue controlled by crack propagation (Miner, 1945). The rule states that in a fatigue test at a constant stress amplitude $\Delta\sigma_i$, damage could be considered to accumulate linearly with the number of cycles. Accordingly, if at a stress amplitude $\Delta\sigma_1$, the component has n_1 cycles of life, which correspond to amount of damage (crack length) a_c, after Δn_1 cycles at a stress amplitude $\Delta\sigma_1$, the amount of damage will be $\frac{\Delta n_1}{n_1} a_c$. After Δn_2 stress cycles spent at a stress amplitude $\Delta\sigma_2$, characterised by a total life of n_2 cycles, the amount of damage will be $\frac{\Delta n_2}{n_2} a_c$ and so on. Failure occurs when, at a certain stress amplitude $\Delta\sigma_M$, the sum of the partial amounts of damage attains the amount a_c, i.e., when

$$\frac{\Delta n_1}{n_1} a_c + \frac{\Delta n_2}{n_2} a_c + \ldots + \frac{\Delta n_M}{n_M} a_c = a_c \tag{14}$$

is fulfilled. As a result, the analytical expression of the Palmgren-Miner rule becomes

$$\sum_{i=1}^{M} \Delta n_i / n_i = 1 \tag{15}$$

where n_i is the number of cycles needed to attain the specified amount of damage (crack length) a_c at a constant stress amplitude $\Delta\sigma_i$.

The Palmgren-Miner rule is central to estimating the fatigue life of components yet no comments are ever made as to whether it is compatible with the rate of damage accumulation law characterising the different stages of fatigue crack growth. The necessary and sufficient condition for the validity of the empirical Palmgren-Miner rule is the possibility to factorise the rate of damage accumulation da/dn as a function of the amount of accumulated damage *a* (the crack length) and as a function of the stress or strain amplitude (Δp):

$$da/dn = F(a)\,G(\Delta p) \tag{16}$$

The theoretical derivation of the Palmgren-Miner rule has been given in (Todinov, 2001). A widely used law for the rate of fatigue damage accumulation is the Paris power law (Paris et al, 1961; Paris & Erdogan, 1963):

$$da(n)/dn = C\Delta K^m \tag{17}$$

where $\Delta K = Y\Delta\sigma\sqrt{\pi a}$, is the stress intensity factor range; *C* and *m* are material constants, *a* is the current crack size (damage) and *Y* is a parameter that can be presented as a function of the amount of damage *a*. Clearly, the Paris-Erdogan rate of damage accumulation law can be factorised as in (16) and therefore it is compatible with the Palmgren-Miner additivity rule. In cases where this factorisation is impossible, the Palmgren-Miner rule does not hold. Such is for example, the fatigue crack growth law

$$da/dn = B\Delta\gamma\, a^{\beta} - D \tag{18}$$

discussed in Miller (1993), which characterises physically small cracks. In equation (18), *B* and β are material constants, $\Delta\gamma$ is the applied shear strain range, '*a*' is the crack size and *D* is a constant.

4.2 Reducing the rate of damage accumulation by derating

Derating is a powerful tool available to the designer for reducing the rate of damage accumulation and the likelihood of failure. It follows directly from the conclusion made in the previous section: *to minimise the total accumulated damage, minimum possible time should be spent at the levels of the damage-inducing factor where the rate of damage accumulation is high.*

Derating to minimize the rate of damage accumulation is done by reducing the operating stresses below their rated levels. Life of many components and systems increases dramatically if the level of the damage-inducing factor is decreased because then, the rate of damage accumulation is decreased.

As can be seen from Fig.6b, reducing the intensity level of the damage-inducing factor from p_1 to p_2 enhances the component's life because of the increased time for the accumulated damage to attain the critical level D_c which precipitates failure. A typical application of this method is the reduction of the stress amplitude, which results in a significant decrease of the rate of fatigue damage accumulation.

Derating essentially 'overdesigns' components by separating the damage resistance capacity distribution from the distribution of the accumulated damage thereby reducing the interaction between the distribution tails (Fig.7). The smaller the interaction of the distribution tails, the smaller the probability that the accumulated damage will exceed the damage resistance capacity, the smaller the probability of failure. In general, the greater the derating the greater the life of the component. However, derating is associated with inefficient use of the damage resistance capacity of components (Fig.7b).

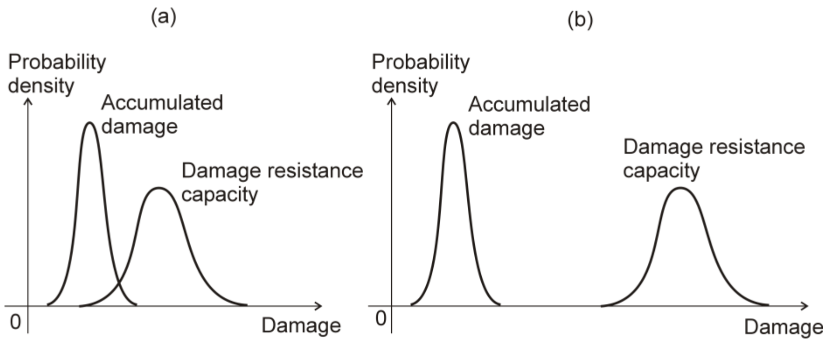

Figure 7. Stress and strength distribution before derating and b) after derating.

Voltage and temperature are common derating stresses for electrical and electronic components. The life of a light bulb designed for 220V, for example, can be enhanced enormously simply by operating it at a voltage below the rated level (e.g. at 110V), which reduces significantly the rate of filament degradation. For mechanical components, common derating stresses reducing the rate of damage accumulation are the operating speed, stress amplitude, temperature, pressure, etc.

4.3 A case related to multiple damage inducing factors

The rate of damage accumulation often depends on multiple damage-inducing factors. Such is the case for the rate dQ/dt of wear debris produced during sliding wear described by the Archard's equation (Archard, 1953):

$$dQ/dt = k\frac{Nv}{H} \quad (19)$$

where k is a constant, N is the normal load, v is the sliding speed, H is the hardness of the softest contact surface and t is the time.

This model states that both the sliding speed and the normal load affect the rate of damage accumulation. The dependence of the damage accumulation rate in the case of mutually independent damage-inducing factors, expressed by equation (19) is very simple. In this case, it is obvious that the rate of damage accumulation is minimized simply by minimising both the normal load N and the sliding speed v and maximizing the hardness. In cases where the damage-inducing factors are mutually dependent (such as temperature and humidity), it is not at all clear what is the optimal variation of the damage-inducing factors yielding the smallest amount of accumulated damage.

In the case of multiple damage-inducing factors varying between a specified initial state and a final state, the optimal trajectory yielding the smallest amount of accumulated damage can be determined by a dynamic programming algorithm (Bellman, 1957). The algorithm will be presented for two damage inducing factors (e.g. temperature and humidity) but it can be easily generalised for any number of damage-inducing factors.

Statement of the problem:

Suppose that there are two damage-inducing factors p and q, whose initial values are p_0, q_0. At the end of the time interval (0,t) the values are p_e, q_e. It is assumed, that at any set of values p^*, q^* for the damage-inducing factors, the time t^* for attaining a critical level of accumulated damage which precipitates failure is specified and can always be evaluated (Fig.8a). Therefore, for a time step Δt, the increase ΔD_p of the damage by varying factor q and keeping factor p constant and the increase ΔD_q of the damage by varying factor p and keeping factor q constant can also be evaluated (Fig.8b).

It is required to determine the optimal trajectory in the space defined by the damage-inducing factors p and q (Fig.8b), such that the total amount of accumulated damage is minimized.

For example, the pair of damage-inducing factors p and q could be: *temperature- humidity*, *concentration-temperature*, *pressure-corrosion potential,* etc.

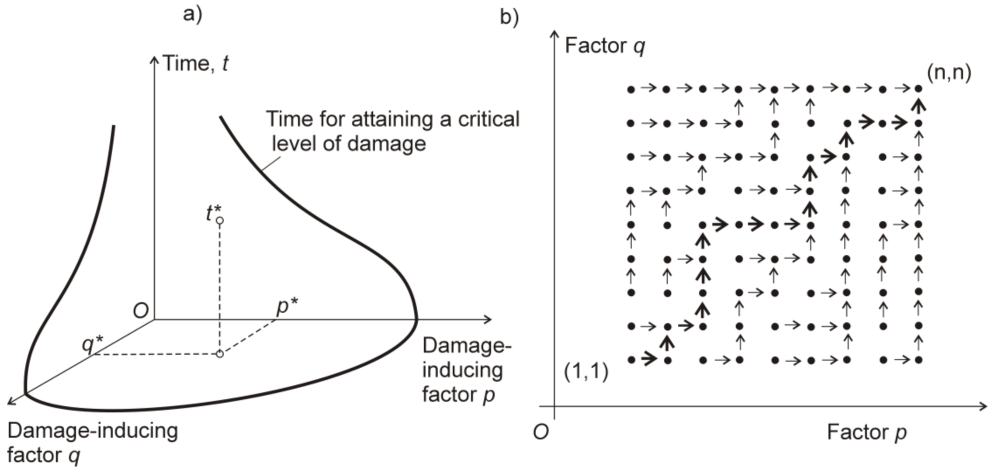

Figure 8. Dynamic table for two damage-inducing factors p and q.

Dynamic algorithm for determining the optimal trajectory yielding the smallest total accumulated damage

The time interval (0,t) during which the damage is accumulated is first divided into 2n steps with length $\Delta t = t/(2n)$ where n is a sufficiently large number to ensure the desired precision. Next, a square dynamic table (two-dimensional array with $n \times n$ entries) is built where n rows are allocated to the different levels of the damage-inducing factor q and n columns are allocated for the different levels of the damage-inducing factor p. It is assumed that the real trajectory in the $p \times q$ phase space is approximated by a staircase-type curve in the $n \times n$ dynamic table, consisting of segments on which one of the damage-inducing factors is constant. This approximation to the real trajectory can be made with any desired precision provided that a sufficiently large number ($2n$) of division points are taken on the time interval $(0,t)$.

The advantage of the dynamic programming algorithm consists of the fact that it finds solutions to sub-problems, stores them in the memory and describes the solution of each sub-problem in terms of already solved and previously stored solutions of other sub-problems. As a result, the sub-problems are solved only once, which makes the dynamic programming algorithm orders of magnitudes more efficient than a brute-force method based on the enumeration of all possible trajectories on the square lattice of size $n \times n$. The number of possible trajectories from node (1,1) to node (n, n) in a square lattice of size $n \times n$ is huge because, for most of the nodes on the square lattice, there are two choices: moving right (increasing p by keeping q constant or moving up (increasing q by keeping p constant). As a result, the computational time of a brute-force method based on scanning all possible trajectories increases exponentially with increasing the number $2n$ of the time intervals into which the interval $(0,t)$ is divided.

The dynamic algorithm for determining the optimal trajectory works as follows. The solutions of the sub-problems are kept in the array x[][], which will be referred to as 'dynamic table'. The size of the x[][] array is $n \times n$. The information necessary to restore the optimal solution is kept in the array trac[][] which is also of size $n \times n$. To each entry of the array trac[n][n] corresponds either the value '1' or the value '2'. The value '1' stands for moving right ('1'), i.e., keeping factor q constant, and the value '2' stands from moving up, i.e. keeping factor p constant.

The last row x[n][] of the array x[][] corresponds to $q = q_e$ value of the parameter q. The only possibility for all entries in the last row is to move 'right', i.e. to keep q constant, therefore the values of the last row of the array trac[][] are all initialized with '1'.

The last column x[][n] of the array x[][] corresponds to $p = p_e$ value of the parameter p. The only possibility for all entries in the last column is to move 'up', i.e. to keep the parameter p constant. Therefore, the values of the last column of the array trac[][] are all initialized with '2'.

The entries $x[i][j]$ of the array x[][] are always initialised with the smallest amount of accumulated damage for moving from node (i,j) to node (n,n).

The entries of the dynamic table (array x[][]) are initialised by starting from entry $x[n][n]$ to which a value zero is assigned: $x[n][n] = 0$. If node $x[n][n-1]$ has been reached, the only direction in the next step is to move right hence, the entry $x[n][n-1]$ is initialised with the value $\Delta D_{(n,n-1),(n,n)}$ which is the smallest amount of accumulated damage from moving from node $(n,n-1)$ to node (n,n). By definition, $\Delta D_{(n,n-1),(n,n)}$ can always be evaluated and is a known quantity. If node $(n,n-2)$, has been reached, the only direction in the next step is again 'right' hence, node $(n,n-2)$ is initialised with the sum $\Delta D_{(n,n-2),(n,n-1)} + x[n][n-1]$ which corresponds to the sum of the smallest amount of damage for moving from node $(n,n-2)$ to node $(n,n-1)$ and the smallest amount of total damage for moving from node $(n,n-1)$ to node (n,n). The entry $x[n][n-2]$ therefore contains the smallest amount of accumulated damage for moving from node $(n,n\text{-}2)$ to node (n, n).

In this way, values are assigned to all entries of the entire last row $x[n][]$. Next, in a similar fashion, values are assigned to all entries of the last column $x[][n]$ of the dynamic table by starting from the entry $x[n\text{-}1,n]$. If an entry from the last column has been reached, the only possibility is to move upwards, therefore, the values from the last column of the array trac[][] are all initialized with '2' which stands for moving 'up'.

Next, the entries of the x[][] array are updated by starting from the entry $x[n-1][n]$. If node $(n-1,n)$ has been reached, the only direction in the next step is upwards and the entry $x[n-1][n]$ is initialised with the value $\Delta D_{(n-1,n),(n,n)}$ which is the smallest amount of damage for moving from node $(n-1,n)$ to node (n,n).

By definition, $\Delta D_{(n-1,n),(n,n)}$ can always be evaluated and is a known quantity. If the trajectory is in the node $(n-2,n)$, the only direction in the next step is 'upwards' and the entry $x[n-2][n]$ is initialised with the sum $\Delta D_{(n-2,n),(n-1,n)} + x[n-1][n]$ which corresponds to the sum of the smallest amount of damage for moving from node $(n-2,n)$ to node $(n-1,n)$ and the smallest amount of accumulated damage for moving from node $(n-1,n)$ to node (n,n). In this way, values are assigned to the elements of the entire last column $x[][n]$ of the dynamic table.

Next, the assignments in the dynamic table continue with node $(n-1,n-1)$. If node $(n-1,n-1)$ has been reached, there are now two possibilities for the next step: to move 'right' or to move 'up'. The decision depends on which quantity is smaller: $\Delta D_{(n-1,n-1),(n-1,n)} + x[n-1][n]$ or $\Delta D_{(n-1,n-1),(n,n-1)} + x[n][n-1]$. If $\Delta D_{(n-1,n-1),(n-1,n)} + x[n-1][n] <$ $\Delta D_{(n-1,n-1),(n,n-1)} + x[n][n-1]$, the trajectory moves 'right', if the converse is true, the trajectory moves 'up'. The entry $x[n-1][n-1]$ is updated with the smaller of the two quantities to show the smallest amount of accumulated damage for moving from node (*n*-1,*n*-1) to node (*n*,*n*). The entry track[n-1][n-1] of the track array is updated with '1' or '2' depending on whether the trajectory moves 'right' or 'up', correspondingly. The process of updating the entries of the *x*[][] array and the track[][] array continues with the remaining entries on the n-1st row starting from entry x[n-2][n-1] and moving towards x[n-2][1]. Next, the row x[n-3][] is updated and this process continues until all entries in the dynamic table are initialised. When the entry x[1][1] is reached, by the way the dynamic table has been constructed, x[1][1] will contain the smallest amount of accumulated damage for moving from node (1, 1) to node (*n*, *n*).

Restoring the optimal trajectory is done by starting from the entry track[1][1]. Initially, *i*=1 and *j*=1. If the recorded value in track[*i*][*j*] is '1', a transition is made to the entry track[*i*+1][j]. If the recorded value in track[*i*][*j*] is '2', the transition is to the entry track[*i*][*j*+1]. The process is repeated until the element track[*n*][*n*] is finally reached. An exemplary optimal trajectory is shown in Fig.8 with bold arrows.

If three damage-inducing factors are present, the dynamic table is a three-dimensional array (x[][][]) and is built in a similar fashion.

The running time of the dynamic algorithm for two damage-inducing factors is therefore $O(n^2)$, for three damage-inducing factors the running time is $O(n^3)$ and so on. Therefore, the dynamic optimisation algorithm runs in polynomial rather than exponential time, which is a big advantage to a brute-force optimisation algorithm.

To minimise the rate of damage accumulation and maximise the life of the component, the trajectory associated with the smallest amount of total accumulated damage must be followed.

5 REDUCING THE RATE OF DAMAGE ACCUMULATION BY DELIBERATE WEAK LINKS

The essence of this mechanism consists of introducing deliberate weak links where damage accumulates instead of accumulating in the valuable part of the system. Instead of accumulating in the expensive component (with a large cost of failure or cost of replacement) the damage accumulates in a cheap component which is replaced when the accumulated damage reaches a critical level. Consequently, the deliberate weak link must be designed with a rate of damage accumulation higher than the rate of damage accumulation of the expensive component.

A good example of this type of deliberate weak links are the sacrificial anodes separating components (underground pipes, underwater installations, ship hulls) from excessive corrosion. Sacrificial anodes form a galvanic couple with the protected component and by corroding preferentially, they protect the expensive component from corrosion. Most commonly, magnesium, zinc and aluminium are used as sacrificial anodes. Sacrificial blocks of magnesium alloy for example are used to protect the steel legs of oil rigs from corrosion. As a result, large supporting metal structures are cheaply protected from corrosion in a highly corrosive sea water environment.

Galvanization (applying protective zinc coating to steel) is most commonly used on outdoor steel structures to prevent rusting. The operation of galvanic anodes is based on the difference in electro-potential and a metal anode and can be used to protect another metal part as long as a sufficient difference in electro-potential is present and the metal anode has a more negative potential than the protected metal part. Thus, components made of copper can be protected by iron anodes because a sufficient difference in electro-potential is present between iron and copper and iron has a more negative potential than copper in the electro-potential series.

Deliberate weak links of this type protect also against failure from excessive wear. For example, cheap inserts in journal bearing take most of the wear and their failure requires the replacement of a cheap insert rather than the replacement of the entire journal bearing.

Spray nozzles are subjected to intensive accumulation of wear damage which requires a frequent replacement of the nozzle. The damage accumulation in the expensive spray nozzle can be avoided by a replaceable nozzle insert. The nozzle insert now takes all the wear damage. Instead of replacing the entire nozzle, only the nozzle insert is replaced.

Rubber segments bolted on top of a metal conveyor belt also act as deliberate weak links reducing the damage accumulation in the expensive conveyor. They take most of the wear and their failure requires the replacement of a cheap rubber segment rather than the replacement of the entire conveyor belt.

For two components in contact, characterised by high contact stresses, one of which is cheap and the other is expensive, the less expensive component should be manufactured from softer material. As a result, the wear and deformation due to the high-magnitude contact stresses will be concentrated in the cheaper component. The consequences of failure will be significantly reduced because the cheap component will be replaced, not the expensive one.

6 REDUCING THE RATE OF DAMAGE ACCUMULATION BY REDUCING EXPOSURE TO ACCELERATION STRESSES

6.1 Reducing exposure to acceleration stresses by reducing the magnitude of the acceleration stresses

The environment is a major source of acceleration stresses which lead to increased rate of damage accumulation and a significant increase of the hazard rates of components. The impact of the environment is manifested through the acceleration stress, which is anything that leads to increased rate of damage accumulation. Examples of acceleration stresses are the temperature, humidity, radiation, vibration, pressure, concentration of particular ions, etc. This list is only a sample of the possible acceleration stresses and can be extended significantly. A typical acceleration stress is the high temperature which increases the rate of damage accumulation and the hazard rates of the affected components. Humidity, corrosion, radiation or vibrations also increase the rate of damage accumulation and the hazard rates of the affected components.

Failure to account for the negative impact of the acceleration stresses usually leads to optimistic reliability predictions - the actual reliability is smaller than the predicted.

Reducing the exposure to acceleration stresses by reducing the magnitude of the acceleration stresses can be illustrated with the most common acceleration stress - the temperature. In many cases, damage is accumulated at elevated temperatures and is caused by the diffusion of particular atoms (ions). As a result, the instantaneous rate of damage accumulation is proportional to the diffusion coefficient of these atoms (ions). Such is, for example, the process of decarburisation of steels which results in a reduced fatigue resistance. The variation of the diffusion coefficient D with temperature, in solids, is given by the equation: $D = D_0 \exp(-E_D / RT)$, where T is the absolute temperature (K), E_D is the activation energy for diffusion (J/mol), R is the gas constant in J/(K x mol) and D_0 is a constant. From this equation, it can be seen that an increase in temperature results in a drastic increase of the coefficient of diffusion and the rate of damage accumulation. Conversely, reducing temperature of the environment drastically reduces the diffusion coefficient and damage accumulation. Limiting the exposure to elevated temperatures therefore, will be of critical importance to reducing the damage accumulation.

6.2 Reducing exposure to acceleration stresses by modifying or replacing working environment

Reducing the exposure to acceleration stresses can also be attained by *modifying or substituting the working environment*.

Arc welding, shielded by an inert gas atmosphere such as argon or carbon dioxide is an example of replacing working environment. As a result, the accumulation of damage from the contact of the weld metal with oxygen is reduced and the reliability of welds is improved. This principle is used, for example, in MIG (metal inert gas) and TIG (tungsten inert gas) welding techniques. Another example is the hermetic or plastic encapsulated integrated electronic circuits to protect them from degradation caused by humidity.

Corrosion increases significantly the rate of damage accumulation. Material degradation due to corrosion (Ohring, 1995) is often the root cause of failures entailing loss of life, damage to the environment and big financial losses. Methods increasing the corrosion resistance include techniques such as *cathodic protection, thermo-chemical treatment of the surface and protective coatings. Corrosion inhibitors* is an alternative way of reducing the rate of damage accumulation from corrosion. These are compounds that modify the corrosive environment thereby reducing the rate of damage from corrosion.

7. REDUCING THE RATE OF DAMAGE ACCUMULATION BY APPROPRIATE MATERIALS SELECTION, DESIGN AND MANUFACTURING

Modifying designs is often used to reduce the rate of damage accumulation. Thus a design modification resulting in a reduced stress range reduces significantly the accumulation of fatigue damage (Todinov, 2016). Zones in components, characterised by large stress gradients within a small volume, are known as '*stress raisers*'. Commonly, these are discontinuities in the geometry and material properties such as *fillets*, *notches*, *holes*, *threads*, *steps*, *grooves*, *keyways*, *rough surface finishes*, *quenching cracks* and *inclusions*.

The stress raisers are reliability-critical design features because they intensify the rate of fatigue damage accumulation. As a result, they make it easy for a crack to initiate and propagate. Consequently, the appropriate design and manufacturing of the stress raisers should be guaranteed. The rate of damage accumulation at stress raisers can be reduced by avoiding sharp notches and corners, keyways, holes, abrupt changes in cross sections, badly machined fillet radii and grooves.

Poor design of the flow paths of fluids containing abrasive material promotes rapid *erosion* which can be minimised by a proper material selection and design. Structural design features promoting rapid erosion should be avoided. The recently introduced domain-independent risk reduction methods: *separation*, *segmentation* and *inversion* (Todinov, 2015) can be used successfully to produce designs associated with a small rate of damage accumulation. Thus, the separation of properties of gears achieved by induction heat treatment results in a hard surface and tough core and yields components with a superior contact strength, fatigue resistance and shock resistance. The segmentation of components contains the spread of damage and prevents the damage accumulation from reaching dangerous levels. Thus, segmenting a pipeline, glass panel, solid rod, etc. confines the spread of a crack within a single segment and limits the amount of accumulated damage.

Introducing inverse states counteracts loading stresses and reduces the intensity of damage accumulation. The *cold expansion*, for example used in aviation for creating compressive residual stresses at the surface of fastener holes is an example of reducing the rate of fatigue damage accumulation by introducing an inverse state. This is done by passing a tapered mandrel through the hole. The compressive residual stress field created in the vicinity of the hole counters the tensile loading stresses during operation and impedes the formation and propagation of fatigue cracks at the edge of the hole.

Material selection is central to design and guarantees the combination of properties needed for the required function. Material selection and selection of appropriate microstructures plays a very important role in reducing the rate of damage accumulation.

Consider the common case of pipes transferring corrosive production fluids or pipes working in corrosive environment. Selecting steel instead of appropriate corrosion-resistant polymer or composite will result in costly failures due to high corrosion rates. In turn, selecting polymer instead of metal or composite for components subjected to intensive cyclic loading often leads to high rates of fatigue damage accumulation and a short fatigue life. In high-temperature aggressive environment, selecting metals instead of ceramics often leads to high rates of damage accumulation and premature failures.

The susceptibility to cavitation damage can be reduced by using cavitation-resistant materials, welded overlay of metals, sprayed metal coatings or elastomeric coatings.

Selecting clean materials, with small size of the flaws, which serve as places for fatigue crack initiation, increases significantly the fatigue life of components. Most of the loading cycles are expended on the early stages of crack propagation when the crack is small. During the late stages of fatigue crack propagation, a relatively small number of cycles is sufficient to extend the crack until failure. The initial flaw size can be decreased by using cleaner material, better material processing and better inspection for flaws.

Controlling the structure of materials during manufacturing, brings unique properties which limit the rate of damage accumulation and the risk of failure in safety-critical applications. Altering the steel microstructure for example, by appropriate heat treatment, often results in a significantly improved fatigue resistance.

The rate of damage accumulation from erosion, for example, is significantly reduced by appropriate heat treatment increasing the surface hardness. The rate of fatigue damage accumulation is reduced significantly by strengthening the components surface by gas carburising, gas nitriding or by the deposition of hard coatings.

Eliminating low-strength surfaces by machining also significantly reduces the rate of fatigue damage accumulation. This is due to eliminating soft decarburized surfaces, eliminating surface discontinuities, folds and pores, and by eliminating coarse microstructure characterized by a low toughness.

CONCLUSIONS

1. The principle of the minimized rate of damage accumulation and its application for improving reliability and reducing risk has been formulated for the first time. Various methods and techniques for reducing the rate of damage accumulation and their application for improving reliability and reducing risk have been presented.

2. Minimizing the rate of damage accumulation has been proposed by a substitution with assemblies working on different physical principles. This method for reducing the rate of damage accumulation has a significant potential yet remains largely unexplored.

3. An optimal replacement for a system with components logically arranged in series has been proposed, resulting in the smallest rate of damage accumulation. It has been proved rigorously that if n spare parts are available for each component, optimal replacement intervals, each of which is $1/(n+1)$ part of the total operational interval, minimize the rate of damage accumulation and maximize the probability of surviving the total operational time interval.

4. A dynamic algorithm has been proposed for the first time for minimizing the rate of damage accumulation by selecting the optimal variation of the damage-inducing factors.

5. A classification of mechanisms and techniques for reducing the rate of damage accumulation has been presented for the first time.

6. The necessary and sufficient condition for using the additivity rule in calculating the damage threshold marking the onset of failure is the factorisation of the rate of damage accumulation law as a function of damage and a function of the damage inducing factor. If the rate of damage accumulation cannot be factorised, the additivity rule for determining the threshold damage which marks the onset of failure is invalid.

REFERENCES

Archard, J.F. (1953). "Contact and Rubbing of Flat Surface". *J. Appl. Phis*. **24** (8): 981–988.

Barlow R.E. and F. Proschan, *Statistical theory of reliability and life testing*, Rinehart and Winston, Inc., New York (1975).

Bellman R., *Dynamic programming*, Princeton University Press, Princeton (1957).

Blischke W.R. and D.N. Murthy, *Reliability: Modelling, prediction, and optimisation*, John Wiley & Sons, Inc., New York (2000).

Childs P.R.N., *Practical Temperature measurement*, Butterworth-Heinemann, Oxford, 2001.

Dasgupta A. and M. Pecht, Material failure mechanisms and damage models, *IEEE Transactions on Reliability*, **40** (5), 531–536 (1991).

French M., *Conceptual design for engineers*, 3rd ed., Springer-Verlag London Ltd, London (1999).

Miller K.J., Materials science perspective of metal fatigue resistance, *Materials Science and Technology*, **9**, 453–462 (1993).

Miner M.A., Cumulative damage in fatigue, *Journal of Applied Mechanics*, **12**, 159–164 (1945).

O'Connor P.D.T, *Practical reliability engineering*, 4th ed., John Wiley & Sons, Ltd., New York (2002).

Ohring M., *Engineering materials science*, Academic Press, Inc., San Diego (1995).

Pecht M., A.Dasgupta, D.Barker, C.T.Leonard. The reliability physics approach to failure prediction modelling, *Quality and Reliability Engineering International*, September/October (4), (1990) 267-273.

Pecht M. Why The Traditional Reliability Prediction Models Do Not Work - Is There An Alternative. *Electronic Cooling* 2(1), (1996) 10-12.

Todinov M.T., Reducing risk through segmentation, permutations, time and space exposure, inverse states and separation, *International Journal of Risk and Contingency Management*, **4** (3), pp.1–21, (2015).

Todinov M.T., *Reliability and risk models: setting reliability requirements*, 2nd ed., Wiley, (2016).

Todinov M.T., Necessary and sufficient condition for additivity in the sense of the Palmgren-Miner rule, *Computational Materials Science*, **21**, 101–110 (2001).

US 3814962 patent, Baermann M., Magnetic worm drive, (1971).